

ASIAN NUCLEAR SAFETY NETWORK

Proposed Structure of the ANSN Web Based Network

PREPARED BY

Hubert Ley
Reactor Analysis and Engineering Division
Argonne National Laboratory
USA

July 19, 2002

CONTENTS

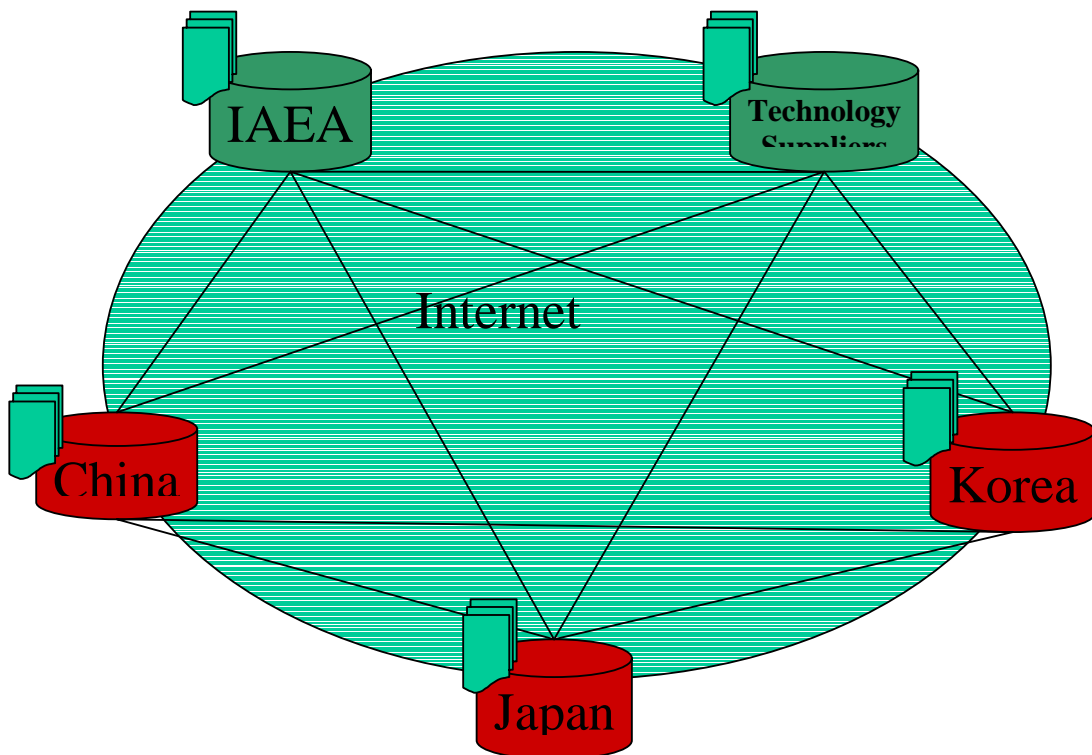
Geographical structure	3
Hub Architecture	3
Hardware and Software	4
Hubs and National Centres	5
Document Formats	5
Indexing and Cataloguing	6
Database Design	6
Relationships of hierarchical tables	7
Attributes of Tables	9
Support Tables	9
Sharing of Index Information	10
Protocols and Standards	10
Pilot Project Limitations	11
Full Development	11
International Network Resources	11

ANSN WEB BASED NETWORK

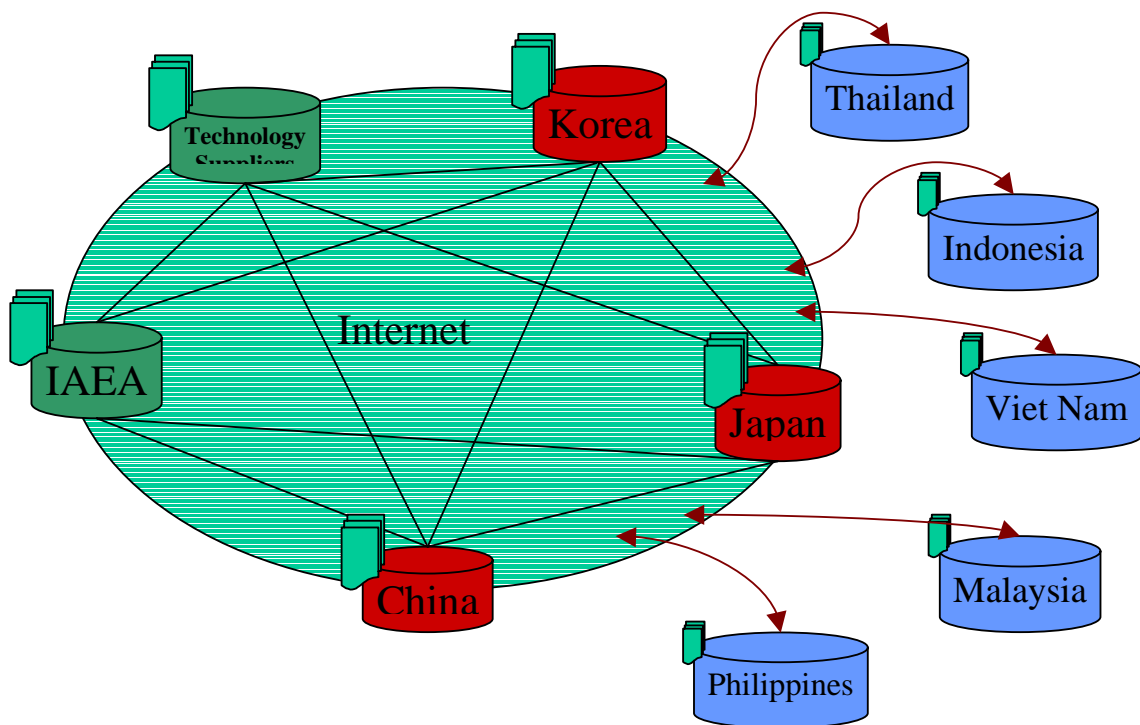
Geographical structure

The major resources will be available from the hubs located in Japan, Korea, and China. Additional resources are expected to be provided by nodes located in Vienna (IAEA), the US, Russia, and possibly Germany and France (this has not been discussed yet). Individual National Centre nodes will be established in Viet Nam, Philippines, Malaysia, Thailand, and Indonesia. Due to the wide spread in the quality of Internet access, some nodes will have significantly better access to the Internet than others. While general improvements of the Internet infrastructure are expected, detailed predictions are difficult to make. For some of the countries, it may be possible to gain access to educational or research network resources. More details on this option will follow later.

Hub Architecture



The diagram shown above shows the general idea of a network of hubs that all have reasonable access to fast Internet service. Each hub is located within organizations that are able to provide content to the proposed ANSN network. Data will be catalogued locally and placed on a web server operated by the local hub. The index resulting from the cataloguing effort will be shared with other ANSN nodes to create a master index that can be used by the end users to locate content of interest.



This diagram shows the National Centre nodes together with hubs shown in the previous figure. It is expected that those nodes have limited Internet access speed, and no node is directly related to a single hub. This limitation is visualized in the graph in form of red arrows connecting the National Centre nodes somewhere to the Internet.

In addition to the hubs and National Centres, individual users of the ANSN will access the Internet from other locations. Users of the ANSN are not included in the graphs to limit the complexity of the figures.

Hardware and Software

Each of the hubs and the National Centres provides a computer for its web site. It also provides a database (preferably located on the same machine) for searching and indexing of the information provided by this node. They all need continuous internet connections to be available as servers for the target end users.

Network security is an important aspect that needs to be provided at each site individually. The exact needs and policies in each country and each location need to be incorporated, but a common minimum standard should emerge from this effort.

Furthermore, each node will provide the necessary trained personnel to operate the hardware, software, and local network. Other personnel will be required to prepare documents (conversion, evaluation), to catalogue the presented information, and to enter the index data into the local database.

Similar equipment being used at each node may reduce the cost of the development and maintenance at each location. Local policies and readily available resources may make the use of existing infrastructure a better option. The choice of resources to be used for the project is one of the first decisions to be made.

Hard disk space may be the most important resource. Other resources may be necessary for the preparation of the data, for example video conversion hardware to create digital copies of analog video, or scanners to turn paper documents into electronic format. Much experience for these conversions is readily available from the US.

For the hubs, network performance is another important factor. The location of the hubs with regards to fast Internet connections is an important aspect.

To clarify the expected needs, a typical expected hardware and software configuration may be a PC with 80 to 160GB of hard disk space (depending on the amount of information provided from the local node), a Linux operating system, Apache web server, PHP scripting language, and a MySQL database. Although all tools mentioned above are free except for the hardware, the configuration and maintenance of the system has a cost associated with the necessary effort. The main reason for the suggested platform is the easy availability of updates for all software, and the fact that all these tools are open software, avoiding dependence on proprietary solutions.

Hubs and National Centres

At this point, the function of the hubs versus the National Centres will be explained in more detail. The hubs are established at organizations and locations that are close to large amounts of relevant information and close to fast Internet connections as well. Information between the hubs can be exchanged quickly and mostly automatically. The National Centres are mostly used to provide services to their local community, and provide access to the information stored on the hubs. While the National Centres are not expected to maintain large amounts of information, they may provide localized information that may not be relevant for the entire ANSN network.

Document Formats

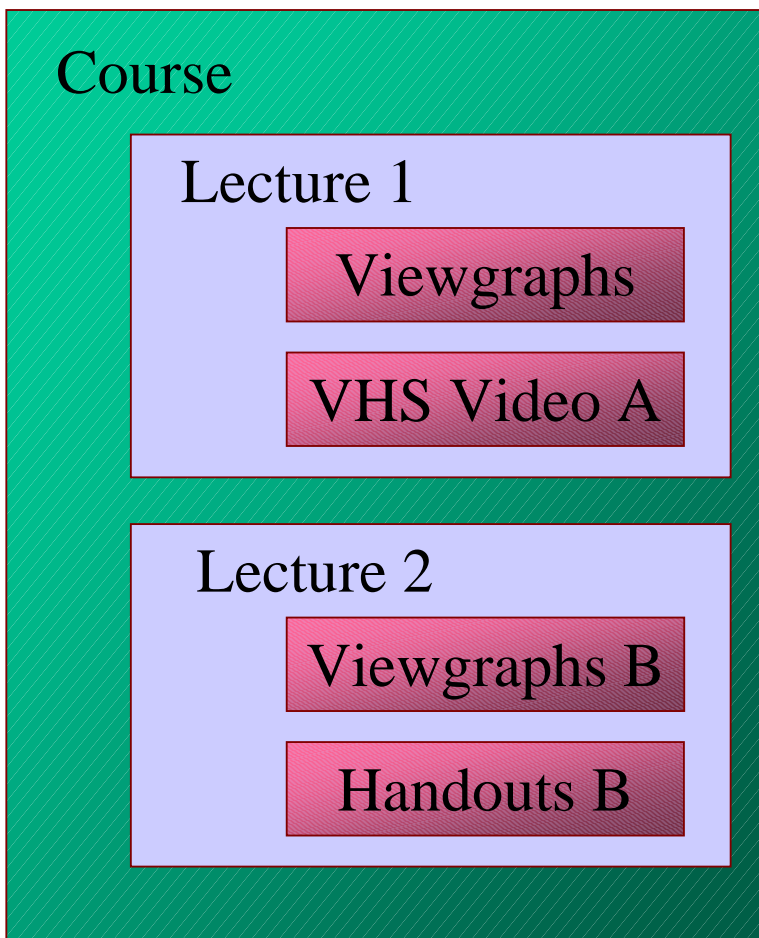
While the emphasis of this project is on providing access to the information in electronic format, it is recognized that this may not be the most effective method in all cases. Some information may be readily available on video tapes or other forms of analog media, and the electronic distribution of the materials may be hampered by large file sizes and long transmission times. The loss of quality for any form of digitization of content may reduce the quality of the material as well. Therefore, information may be catalogued and made searchable in the index without being available in electronic format. An ordering mechanism needs to be established to allow for the distribution of such materials.

Indexing and Cataloguing

Information available to each site is catalogued according to a standardized format. This format will be outlined in another presentation and is not subject of this document. The resulting local index is combined with the corresponding entries from the other nodes to create a complete master index. The exact method for this combination into a master index will be described later. The master index is provided to the users of each local web site node, allowing searching, browsing, and possible downloading or ordering of all available materials. This method ensures that a user of the ANSN network can always search the complete network for information, not just the local node the user connected to.

Database Design

The database design is based on a hierarchical structure. This structure is necessary to implement a meaningful index of the available information, but will be fully transparent to the end user. A typical collection of documents is shown below:



The need for hierarchical grouping becomes obvious. A safety course (or other types of documents) may consist of one or more logical units, which in turn may consist of a number of media formats (handouts, viewgraphs).

Based on this need, all information stored in the database consists of entries on three different levels:

- Lowest level: individual items such as files (ITEM).
- Intermediate level: groups of one or more items creating a logical unit, such as a lecture (DOCUMENT).
- Highest level: groups of one or more lectures, such as a nuclear safety course (GROUP).

The level on which an end user would typically access documents is probably the DOCUMENT level. Under certain circumstances, it may be preferable to access documents by their GROUP or by criteria based on individual ITEMS. The path a potential end user takes to access any of the documents is programmed later in the web site interface, and is not an explicit aspect in the database design. The database design should rather accommodate any number of useful ways to search the database, or to allow access by category, keyword, author, organization, and so on.

An end user would need no (or only a limited) understanding of the three table levels. He would need to understand the basic concept that a document can be grouped into larger units, that it can be part of more than one group (e.g. a class that is a component of several courses), and that a document can exist of a number of physical items, such as paper, video, and other media.

Relationships of hierarchical tables

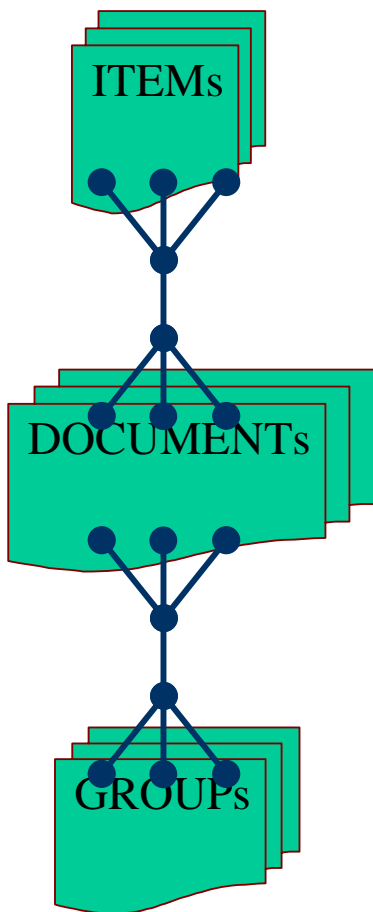
The personnel cataloguing the materials need a more thorough understanding of this tabular concept. The definitions would be slightly refined:

- ITEM
 - Represents the smallest inseparable physical units that can be stored and/or transmitted
 - Can be part of one or more DOCUMENTs
- DOCUMENT
 - Consists of one or more ITEMS that represent the smallest logical unit
 - Can be part of one or more GROUPs
- GROUP
 - Consists of a number of DOCUMENTs forming a logical volume

The important fact is the type of relationship between the tables. First of all, there are three basic types of relationships possible between tables with related content:

- “One to one” relationship – each entry in one table has only one corresponding entry in the second table, and vice versa. Both tables would have an identical number of rows.
- “One to many” relationship – a single entry in one table may have more than one corresponding entry in the other table, but not the other way around. Example: people on a car trip: Any given person can be only in one car at a time, but one car can transport more than one person.
- “Many to many” relationship – a single entry in one table can have more than one corresponding entry in the other tables, and vice versa. Example: groups of people: Each person can belong to more than one group, and a group consists typically of more than one person.

All three tables (ITEM, DOCUMENT, GROUP) are hierarchically related through “many to many” relationships. This leads to the following diagram:



representation.

Any item can be part of one or more documents, and a document consists of one or more items.

Any document can be part of one or more groups, and a group consists of one or more documents.

There are also some additional variations of the concept: to allow for more flexibility, relationships do not need to be defined in all cases. While a document needs certainly to contain at least one item, it does not have to belong to any group. As a matter of fact, this case may be the most common case. In a database, the non-existence of a link between hierarchical tables is perfectly allowable.

The same is true for items: There may be special cases where items are entered into the database without belonging to any specific document. In this case, the database could allow for such cases. While designing the database, the designer has to use his judgment on whether to allow such missing relationships or not.

Without going further into the details of database development, it should be noted that the database designer has to balance restrictions against flexibility to allow for the most organized approach to data search and evaluation, while not restricting data entry and

Attributes of Tables

When adding material to the database, the material has to be catalogued. A number of attributes exist that describe the material in the database. This is a critical point, because important decisions have to be made. First, is this a single document or does it belong to one or more larger groups of documents? This defines the hierarchical relationship between document and group.

Then it has to be determined whether there are more than one physical item. For electronic documents, they could exist in equivalent word processor and PDF files, or of a collection of more than one file in the same or different formats. In this case, the collection of files could be packaged into a single archive file (ZIP), and they would be treated furthermore as a single item. This would not work for analog items such as video tapes. Nevertheless, a digitization and subsequent creation of a CDROM with all materials would also create a single item. Packaging physical items in this manner may reduce the cataloguing effort to some degree, at the cost of additional burden and download times for the end user. This is a good example how more detailed cataloguing effort can improve the end user experience. The decision on how to catalogue the information is therefore crucial for the whole process.

Besides the breakdown into items, documents, and groups, attributes have to be provided for each of the three categories. On the GROUP level, there are probably only few attributes, like Title, Organization, Author (or Contact Person), Date, and a description.

The DOCUMENT level is probably the level with the largest number of attributes, such as: Title, Organization, Author, Date, Abstract, Category, Technical Area, Level of Expertise, Area of Competence, Prerequisites, Learning Mechanism, References, and Keywords.

Some attributes apply on the ITEM level: Again a Title, Author, and Date. Additionally, there is a need for a short Description, Format and Medium, Language, Filename or Identification, Equipment Needed, Special Considerations, and possibly Keywords.

The decision on the attributes on each level is not finished. Future discussion will determine the minimum set of useful attributes. Empty attributes must be allowed for individual entries. Too many attributes will make the data entry more costly.

Support Tables

A number of support tables is necessary to simplify data entry and to avoid double entries. They consist at least of the following set of tables:

Countries, Organizations, Individuals, Organization Membership (linking individuals to organizations), Technical Areas, Level of Expertise, Area of Competence, Format and Medium, and Language.

Each of these tables provides a limited number of entries, that can be use either in form of pull-down menus or other indexing methods.

Sharing of Index Information

The previous chapters defined the data entry and index creation for an individual ANSN node. To create an integrated user experience, the local indexes on each QANSN node need to be integrated into a single master index that allows a complete search or browsing of the whole collection of documents.

Three basic methods available

- Each site maintains and uses only their own index (least desirable solution)
- A master site is chosen with a copy of all index information
- Each site receives read-only copies of all other site indexes and assembles a full list together with the locally maintained index (fault tolerant solution)

End users need to be informed of added entries, e.g. by E-mail or a list of recent postings

At least for the pilot projects, it seems advisable to opt for a single master index located on one of the ANSN nodes. While each node prepares and maintains their individual index of the information stored locally, the master index is updated on a regular basis and should be used by all ANSN nodes as the main entry point for the end user.

A future development may be the distributed concept of master index information. While no single master index exists in this concept, every node receives the index information from all other nodes as read-only copies. Each node is then able to create its own version of the master index by combining their own local index with the read-only copies of all other nodes. This approach leads to a more fault proof system, but is more difficult to establish. The distributed index copy may be established at a later time if determined to be necessary.

Also, the actual information (electronic files) will remain as a single copy on each local node (mostly on the hubs). If a document is found by an end user in the master index, he will still download the material from the node that provided this information in the first place. For performance reasons, this might also be upgraded to a distributed concept, where the data is available from more than a single node (for example from all hubs). Due to the increased complexity of the system and the undetermined need for such a solution, this will be considered an option for the future.

Protocols and Standards

To simplify the distribution of documents, the system of ANSN nodes should make use of a number of easily available standards and methodologies. Data transmission should use the TCP/IP protocol (basically, the web) and avoid other proprietary standards. Web sites should make use of existing HTTP, HTML, and XML standards. Extensions to those

standards may be used if widely used and easily available for implementation. This includes scripting environments (PERL, PHP, etc.).

A standard database system should be used, based on SQL and connectable by ODBC drivers. Data exchange may also be implemented by using the extensible markup language XML.

The data cataloguing standards are developed separately and are subject of a different presentation.

Pilot Project Limitations

- Start with a limited number of hubs
- Implementation of single master index only
- Proof of feasibility of implementation
- Improved efficiency of initial data entry:
- Grouping documents into larger units allows to reduce the initial cataloguing effort
- Later replacement with more finely grained catalogue entries will improve search and download capabilities

Full Development

- Implementation of all ANSN nodes
- Integration of standard document access into locally adapted web sites
- Improvement of database and user interfaces based on results of the pilot project
- Detailed testing
- Addition of all documents intended to be distributed through the ANSN Network
- Expansion into knowledge management

International Network Resources

The ANSN may be able to obtain improved network access by tapping into existing network resources, such as

- STAR TAP - The Science, Technology, and Research Transit Access Point (US)
- APAN- Asia-Pacific Advanced Network
- CERNET- China Education and Research Network
- KOREN - Korean advanced Research Network
- KREONet2 - Korea Research Environment Open Network 2
- SingAREN - Singapore Advanced Research and Education Network
- TransPAC - connection of the Asia Pacific Advanced Network (APAN) to STAR TAP. APAN members include Australia, Japan, Korea and Singapore
- SINET - Japan's Science Information Network
- UniNet - Thailand's Education and Research Network

These networks are typically operated by major research and educational centres. Representative from each country may be able to gain access to those networks for the ANSN nodes. The speed on these networks allows for advanced communication, such as Internet Video Conferencing, allowing closer cooperation between member countries.